

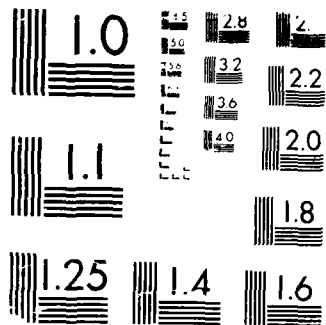
AD-A194 564 THE IMPACT OF USING TRUNCATED DATA TO ESTIMATE VARIANCE 1/1
(U) JET PROPULSION LAB PASADENA CA H W RENNIE
06 JAN 87 JPL-D-4620 NAS7-918

UNCLASSIFIED

F/G 12/3

NL





MICROCOPY RESOLUTION TEST CHART

1963-A

AD-A194 564

DTIC FILE COPY

2

7057-68

U.S. ARMY INTELLIGENCE CENTER AND SCHOOL
SOFTWARE ANALYSIS AND MANAGEMENT SYSTEM

THE IMPACT OF USING TRUNCATED DATA TO ESTIMATE VARIANCE

TECHNICAL MEMORANDUM No. 23

MARC

Mathematical Analysis Research Corporation



06 January 1987

National Aeronautics and
Space Administration

JPL

JET PROPULSION LABORATORY
California Institute of Technology
Pasadena, California

JPL D-4620
ALGO_PUB_0093

DTIC
ELECTE
MAY 26 1988
S H D

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

88 5 26 03 4

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ALGO_PUB_0093	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Technical Memo 23, "The Impact of Using Truncated Data to Estimate Variance"		5. TYPE OF REPORT & PERIOD COVERED FINAL
		6. PERFORMING ORG. REPORT NUMBER I-4620
7. AUTHOR(s) Mathematical Analysis Research Corp. (MARC)		8. CONTRACT OR GRANT NUMBER(s) NAS7-018
9. PERFORMING ORGANIZATION NAME AND ADDRESS Jet Propulsion Laboratory, ATTN: 171-209 California Institute of Technology 4800 Oak Grove, Pasadena, CA 91109		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS RE 182 AMEND #187
11. CONTROLLING OFFICE NAME AND ADDRESS Commander, USAICS ATTN: ATSI-CD-SF Ft. Huachuca, AZ 85613-7000		12. REPORT DATE 06 Jan 87
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Commander, USAICS ATTN: ATSI-CD-SF Ft. Huachuca, AZ 85613-7000		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE NONE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Dissemination		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Prepared by Jet Propulsion Laboratory for the US Army Intelligence Center and School's Combat Developer's Support Facility.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Fix Estimation, Angular Variance, Truncated Data, Cutoff Value, Simulation, Error Ellipse, QUICKFIX		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report examines the effect of truncation of data on the estimation of variance in fix estimation. The methodology used in QUICKFIX is examined in detail. The effect of different truncation parameters and different numbers of LOEs are tabulated. Under the assumption that truncation is centered about the true mean, a formula for correcting for truncation is presented in Appendix I. Appendix II is missing and can be obtained by contacting MARC directly.		

7057-68


U.S. ARMY INTELLIGENCE CENTER AND SCHOOL
Software Analysis and Management System

The Impact Of Using Truncated
Data To Estimate Variance

Technical Memorandum No. 23


06 January 1987

Author:

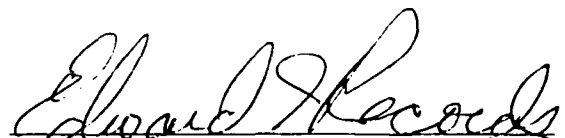


MARC
Mathematical Analysis Research Corporation

Approval:

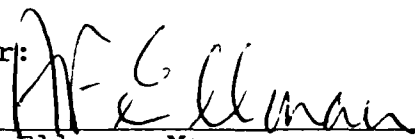


James W. Gillis, Subgroup Leader
Algorithm Analysis Subgroup




Edward J. Records, Supervisor
USAMS Task

Concur:



A. F. Ellman, Manager
Ground Data Systems Section



Fred Vote, Manager
Advanced Tactical Systems

JET PROPULSION LABORATORY
California Institute of Technology
Pasadena, California

JPL D-4620

PREFACE

The work described in this publication was performed by the Mathematical Analysis Research Corporation (MARC) under contract to the Jet Propulsion Laboratory, an operating division of the California Institute of Technology. This activity is sponsored by the Jet Propulsion Laboratory under contract NAS7-918, RE182, A187 with the National Aeronautics and Space Administration, for the United States Army Intelligence Center and School.

This specific work was performed in accordance with the FY-87 statement of work (SOW #2).

Accession For	
Dist	<input checked="" type="checkbox"/>
Dist	<input type="checkbox"/>
Dist	<input type="checkbox"/>
Availability Codes	
Dist	Special
A-1	

The Impact of Using Truncated Data To Estimate Variance

The impact of using truncated data to estimate variance is that one's estimates are biased small. This results in error ellipses that are too small. The amount of this bias depends on how close the true standard deviation (=square root of the variance) is to the truncation point. It can be insignificant or very significant. What about real systems?

Guardrail V: An angular truncation value of 4.5 degrees is used in determining which data is used. The fix method used is a 3-dimensional weighted perpendicular. An estimated covariance is used with an F cutoff. Review of this method did not raise any major concerns because MARC was led to believe that the true standard deviation of incoming data would be much smaller than 4.5 degrees.

Quickfix: An angular truncation value of 4 degrees (possibly 6 was intended as there is a contradictory comment in the computer code used in analysis of the algorithm) determines which data is used. The fix method used is a 2-dimensional weighted perpendicular. A standard deviation of 3 degrees is used unless the estimated standard deviation is bigger. If it is bigger then the estimated standard deviation is used. Thus the estimated standard deviation is only used when it is close to the truncation point.

Analysis of Quickfix suggests a need to understand how truncation affects estimates of covariance (and hence standard deviation). A full analysis would require a bigger study than will be performed here, nonetheless the basic impact of truncation is can be understood in terms of the following observations:

A) The existence of code for the large sigma case in Quickfix suggests the possibility that either

1) large sigma are possible and the truncation value can be close to the cutoff value

or 2) the assumption of normality is suspected and the code was designed to cover the "fat tails" problem. The "fat tails" problem occurs when the probability of a large error is higher than predicted by the normal.

B) One would seldom invoke the special code in Quickfix except in the small sample size case. (The small sample size case invokes it more frequently both because the estimator for variance is computed dividing

by $n-3$ instead of $n-2$ and because variance of the estimate of variance calculation is greater in this case.) Representative cases are computed in Tables I and II below.

- C) Truncated normals produce low estimates of variance. See table III. The division by $n-3$ in the Quickfix variance calculation instead of $n-2$ has a corrective impact in the small sample size case. See Tables IV and V.

CALCULATIONS

The calculations are meant to be suggestive. Simplifying assumptions have been made to allow this study to be performed without interfering with other work. In Tables I, III, IV and V the primary simplifying assumption is that the truncation is centered about the true mean. This assumption is really only acceptable for the large sample size case. A formula for recomputing tables such as Tables III, IV, and V may be found in Appendix I. The reason for not including these here is that MARC does not have enough background on the cases of interest to provide narrow down the cases computed to a small enough set for inclusion. Similarly Table I may be recomputed for with truncation not centered about the mean by adding a constant to X in the computer program attached in Appendix II.

TABLE I
Percentage of Quickfix Estimates That Are Greater
Than Three Degrees Assuming An Underlying Normal

Number of Lines of Bearing Used In the Estimate	True Standard Deviation	Percentage of Estimates Greater
4	3.5	83%
5	3.5	64%
6	3.5	47%
7	3.5	33%
8	3.5	23%
9	3.5	15%
10	3.5	10%
4	4.0	85%
5	4.0	67%
6	4.0	51%
7	4.0	37%
8	4.0	27%
9	4.0	19%
10	4.0	13%
4	5.0	87%
5	5.0	71%
6	5.0	56%
7	5.0	43%
8	5.0	31%
9	5.0	23%
10	5.0	17%

TABLE II
Percentage of Quickfix Estimates That Are Greater
Than Three Degrees Assuming Uniform Error

Number of Lines of Bearing Used In the Estimate	Percentage of Estimates Greater
4	90%
5	78%
6	65%
7	53%
8	42%
9	33%
10	26%

Note that uniform may be considered to be normal with very high sigma or wide tailed normal exaggerated. Furthermore centering at the true mean does not matter. Uniform is probably a bound on the optimistic side.

Tables III, IV and V which follow overestimate the impact somewhat as they assume the truncation is centered about the true mean. Since the true mean is not known this assumption is not warranted. An asymmetric distribution expression is derived in the appendix that can be used to determine sensitivity to lack of knowledge of the true mean.

TABLE III
Reduction In Expected Variance Because of Truncation
GENERAL CASE

Truncation Point In Standard Deviations	Expected Value of the Computed Variance As A Percentage Of The True Variance
3.5	.9939
3.0	.9733
2.5	.9113
2.0	.7738
1.5	.5515
1.0	.2910
0.5	.0808

TABLE IV
Reduction In EXPECTED Variance Because of Truncation
SPECIAL CASE-QUICKFIX Truncation At 4 Degrees

True Standard Deviation of Angular Error	Expected Value of the Computed Variance As A Percentage Of The True Variance
10 Degrees	.0521
5 Degrees	.1956
4 Degrees	.2910

QUICKFIX does not use the standard variance formula however. Where one normally divides by $n-2$ where n is sample size, QUICKFIX appears to divide by $n-3$. It is unclear if the use of $n-3$ was an attempt to compensate for truncation. In any case using this adjustment one can compute values such as those in Table V that show the net effect of truncation and dividing by the wrong normalization constant, i.e. $n-3$ instead of $n-2$.

TABLE V
Reduction In COMPUTED Variance Because of Truncation
SPECIAL CASE-QUICKFIX Truncation At 4 Degrees

True Standard Deviation of Angular Error	Expected Value of the 'QUICKFIX' Variance As A Percentage Of The True Variance			
	n=4	n=5	...	n=∞
10 Degrees	.1141	.0781		.0521
5 Degrees	.3912	.2934		.1956
4 Degrees	.5821	.4365		.2910

For small sample size the fact that the variance is not being computed about the true mean would also raise these values somewhat. For small sample sizes and standard deviations close to 3 degrees, the fact that 3 degrees is the minimum allowed would increase the percentage as well. A complete analysis of the ways these factors interact would require simulation. There does not appear to be any reason to increase the sample size=4 case more than for other sample sizes, however. Thus it is our opinion that adjusting all sample sizes the same amount would be preferable to QUICKFIX's approach of dividing by n-3.

APPENDIX I

Truncated Estimates of Variance

Without loss of generality assume $X \sim N(0, \sigma^2)$.

For ease of notation assume $N(x)$ denotes the value of the cumulative density of the $N(0,1)$ normal at x .

Then $\text{VAR}_{\text{Truncated}}(X) = E_{\text{Truncated}}(X^2)$

$$= \frac{\int_a^b x^2 \exp(-x^2/(2\sigma^2)) dx}{\int_a^b \exp(-x^2/(2\sigma^2)) dx}$$

$$= \sigma^2 \{1 - [1/\sqrt{2\pi}] [(b/\sigma) \exp(-(b/\sigma)^2/2)) - (a/\sigma) \exp(-(a/\sigma)^2/2))\} / [N(b/\sigma) - N(a/\sigma)]$$

The table shown in this memo is for $a=-b$.

END

DATE

FILMED

8-88

DTIC